

# Effectiveness of Variables Used in the Model to Detect Discrepant Results During Reinterview and the Identification of New Variables

## FINAL REPORT

This evaluation reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

---

Carrie Johanson  
Decennial Statistical  
Studies Division



# CONTENTS

|   |       |
|---|-------|
| EXECUTIVE SUMMARY .....   | iv    |
| 1. BACKGROUND .....   | 1     |
| 1.1 What were the components to the Census 2000 reinterview program? .....  | 1     |
| 1.2 How was Census 2000 reinterview conducted? .....  | 2     |
| 1.3 How were data about Census 2000 reinterview selection and results obtained for the evaluation? .....                                  | 2     |
| 2. METHODS .....  | 3     |
| 2.1 How were forms identified for data capture? .....   | 3     |
| 2.2 How were cases weighted to account for data capture? .....  | 3     |
| 2.3 How were enumerator summaries made? .....   | 4     |
| 2.4 What analysis techniques were used? .....   | 5     |
| 3. LIMITS .....   | 5     |
| 3.1 Limitations to evaluation responses for all of the research questions .....   | 5     |
| 3.2 Data quality limitations .....  | 6     |
| 3.3 Reinterview program limitations .....   | 6     |
| 4. RESULTS .....  | 7     |
| 4.1 What was the Census 2000 NRFU reinterview workload? .....   | 7     |
| 4.2 How many enumerators were selected for reinterview? .....   | 7     |
| 4.3 How many enumerators were identified as outliers by the administrative test? .....  | 9     |
| 4.4 What percent of the enumerators identified as outliers by the administrative test were selected for administrative reinterview? ..... | 10    |
| 4.5 What percent of the selected reinterview cases contain useable results? .....   | 11    |
| 4.6 What percent of the completed reinterview cases contained discrepant results? ....  | 11    |
| 4.7 What percent of enumerators with complete reinterview results had any cases with errors? .....  | 12    |
| 4.8 How effective was the administrative reinterview in identifying enumerators with errors? .....  | 13    |
| 4.9 What is the contribution of each of the Census 2000 variables in identifying cases with discrepant results? .....                     | 15    |
| 4.10 How do training and personal characteristics of the enumerators with discrepancies compare to all of the NRFU enumerators? .....     | 15    |
| 4.11 Could any of these new variables be used to predict or target enumerators who would be more likely to have discrepant results? ..... | 17    |
| 5. CONCLUSIONS AND RECOMMENDATIONS .....  | 17    |
| REFERENCES .....  | 18    |
| APPENDIX: TECHNICAL DOCUMENTATION .....   | A - 1 |

## List of Tables

|  |    |
|--|----|
| Table 4.1.1 Totals and percent of cases selected for reinterview by sample type . . . . .  | 7  |
| Table 4.2.1 Enumerators selected for reinterview by sample type, categorized by total number of forms selected for reinterview for the enumerator . . . . .                      | 8  |
| Table 4.3.1 Number of times each of the D908 variables contributed to an enumerator being an outlier and the number of unique outlier enumerators for each variable . . . . .    | 9  |
| Table 4.5.1 Percent of selected reinterview cases with complete reinterview, by sample type .  | 11 |
| Table 4.6.1 Error rates for completed reinterview cases by sample type and type of error . . . .   | 12 |
| Table 4.7.1 Percentage of enumerators with completed reinterview cases by categories of error  | 13 |
| Table 4.8.1 Percent of cases contributing to enumerators with error cases, by sample type and number of forms selected for reinterview . . . . .                                 | 14 |
| Table 4.8.2 Frequencies of administrative enumerators across steps of reinterview and data capture process . . . . .   | 14 |
| Table 4.9.1 Percentages of reinterviewed enumerators identified as outliers, by variable in the administrative test and with cases in error (enumerator level dataset) . . . . . | 15 |
| Table 4.10.1 New variables descriptions and categories . . . . .   | 16 |
| Table 4.10.2 Distribution of all NRFU enumerators and enumerators with errors for each new variable . . . . .  | 16 |

## Appendix

|  |       |
|--|-------|
| Form D-908 (NRFU)Administrative Reinterview Trouble Report .....                           | A-2   |
| Form D-191 (NRFU-RI) Reinterview Control Record .....                                      | A-3   |
| Form D-806 Reinterview and Reconciliation Questionnaire .....                              | A-4   |
| Table A Unduplicated case level counts through the reinterview process by case sample type | A-8   |
| Table B Reinterview case sample types by enumerator .....                                  | A-12  |
| Table C Results from discriminant regression analysis .....                                | A- 13 |

## **Executive Summary**

The Census 2000 Nonresponse Followup Reinterview program included three components: a random reinterview, an administrative reinterview, and a supplemental reinterview. A portion of completed enumerator questionnaires were selected to be reinterviewed, and once the reinterview was conducted, the unit status and household roster were compared to the original enumeration. The purpose of the reinterview program was to identify faulty data collection, both intentional and unintentional.

This evaluation looks at the effectiveness of the administrative reinterview and the contributions of the characteristics, or variables, used to identify enumerators for administrative reinterview.

The random reinterview component was designed to verify work from each enumerator. Virtually all enumerators who completed a minimum of ten enumerator questionnaires had one or more of their questionnaires selected for random reinterview. Random reinterviews represented 93.09 percent of the cases selected for the reinterview program. The remainder of the reinterview cases were administrative and supplemental reinterview cases (4.34 percent and 2.57 percent, respectively).

Outlier enumerators were identified for administrative reinterview by comparing questionnaire characteristics of each enumerator against the average for their area. A high vacancy rate, a high rate of partial interviews, a high delete rate, a high rate of questionnaires with a population count of one, and differences in average population per household were variables used in the comparison. The reports identifying these outlier enumerators were run once a week. Over the entire Nonresponse Followup Operation, 291,441 enumerators were flagged as outliers for one of the reasons above. This is 62.57 percent of enumerators with completed work. Not all of these enumerators had cases selected for administrative reinterview. At the discretion of supervisors, approximately five percent of enumerators flagged for administrative reinterview had administrative cases selected, or 3.5 percent of all Nonresponse Followup enumerators.

Supplemental reinterview could be used any time there was reason to suspect cases might not be completed correctly. Supplemental cases with complete reinterview information show a higher frequency of enumerator error between the original enumeration and the reinterview (11.30 percent) than random and administrative cases (9.42 percent and 9.67 percent, respectively). This higher incidence of error identification shows the effectiveness of the supplemental reinterview component.

At the individual case level, administrative and random reinterview found a similar proportion of cases with discrepancies. Of the enumerators in administrative reinterview, 52.09 percent had one or more cases in error. This is much higher than the percent of enumerators in random reinterview with error cases. The range of error rates for enumerators, depending on the number of random reinterview cases selected, was approximately 10 to 14 percent.

We ran a discriminant regression model to evaluate the effectiveness of various variables in the administrative reinterview. We regressed the presence of an error against the variables used for the Census 2000 administrative reinterview and some new variables. The new variables were enumerator-level characteristics: hours in training, education level, test score, any additional language ability, and previous enumeration experience. We were looking for interactions that might help us improve the model we use to identify outlier enumerators for the administrative reinterview.

Of the characteristics reviewed for the administrative sample, the high delete variable had the biggest impact for identifying enumerators with error. However, our regression models showed that very little of the variance associated with the dependent variable (presence of error) was explained by the independent variables of interest. This indicates that we could expect the dependent variable to behave similarly for randomly selected enumerators and enumerators identified as outliers.

Interpreting these results is difficult because of operational limitations. Our analysis shows that administrative reinterview was definitely effective in identifying enumerators with error, yet the contribution of the variables we used to select the enumerators was not meaningful. This is partially explained by the fact that although we targeted enumerators based on work characteristics, our selection of cases for reinterview did not reflect the characteristic(s) that caused the enumerators' outlier status. Other studies and academic experts have shown that a targeted reinterview can be very beneficial. An example of targeted reinterview would be if an enumerator was an outlier for high deletes, then deleted cases were specifically reinterviewed. Using targeted case selection will likely make the administrative reinterview more effective.

We recommend reducing the number of enumerators identified as outliers. This can be accomplished by increasing the critical cut-off levels or accounting for multiple tests and the recurring time periods. Reducing the outliers identified by the administrative test will create a smaller workload to review. This, in turn, should increase the percentage of enumerators reinterviewed, and the effectiveness of the administrative reinterview program in identifying enumerators with discrepant results.

## **1. BACKGROUND**

### **1.1 What were the components to the Census 2000 reinterview program?**

The Census 2000 Quality Assurance (QA) programs for the Nonresponse Followup (NRFU) operation included reinterviews of selected cases. The objective of the reinterview program was to identify faulty data collection, both intentional and unintentional. We achieved this objective through the combination of an administrative sample, a randomly selected sample of each enumerator's questionnaires, and/or a supplemental reinterview.

The Local Census Office (LCO) reinterview staff reinterviewed a selection of seven random cases from each enumerator's workload. This was an automated selection process. One of the first ten cases an enumerator completed was randomly selected. Systematically, another six were selected from the enumerator's workload until the random batch was completed. So, if fewer than seven cases were selected for an enumerator, the enumerator should have completed fewer than 70 questionnaires. The random reinterview was designed to identify enumerators not producing quality work early in the operation, to minimize faulty data collection and rework.

If an enumerator's work characteristics were out of tolerance when compared to the work characteristics of the other enumerators in the same work area, then that enumerator was identified as an outlier on an Administrative Reinterview Trouble Report (D908). The work characteristics reviewed were:

- a. Average population per household
- b. Vacancy Rate
- c. Partial Interview Rate
- d. Delete Rate
- e. Population Count of One Rate

The D908 report was reviewed by the Office Operations Supervisor (OOS) for Reinterview in the office. Unless a justification was given, enumerators were to be identified for administrative reinterview in the tracking database. The next ten cases submitted by that enumerator were selected for reinterview. The administrative reinterview was designed to use available information about the enumerator and their work characteristics to identify faulty data collection.

Supplemental reinterview could be used at any time. Crew Leaders used their discretion to identify enumerators for supplemental reinterview. A large number of forms with unknown telephone numbers, a lot of work completed in a short amount of time, or forms returned without the appearance of normal wear and tear may be reasons to select enumerators for supplemental reinterview. Once an enumerator was identified in the system for supplemental reinterview, the next ten cases submitted by that enumerator were selected for reinterview.



Each case was assigned a sample type of random, administrative, or supplemental. Each batch was also assigned a batch sample type of random, administrative, or supplemental. Generally a random batch contained all random selection cases and an administrative batch all administrative selected cases. A batch could have been started as a random batch and then converted to an administrative batch. In those situations an administrative batch contained both random and administrative cases.

## **1.2 How was Census 2000 reinterview conducted?**

Once a case, or enumerator questionnaire, was selected for reinterview, a label was printed for a Reinterview and Reconciliation Questionnaire (D806). Both the enumerator questionnaire and the D806 were moved to a transcription area, where office clerks transcribed some basic information (e.g. address, unit status, householder names) from the enumerator questionnaire to the D806 form. Telephone reinterviewers made up to six attempts to contact the household respondent by telephone to conduct the reinterview. When a telephone reinterview was not successful, up to three personal visits were attempted by reinterview enumerators in an attempt to conduct the reinterview.

During the reinterview, reinterviewers collected unit status and householder names. The reinterview staff would then make a decision by comparing the original enumeration data to the reinterview data. Reinterview enumerators were to classify cases as 'reject' when the housing unit status collected in reinterview did not match the original unit status and/or at least 50 percent of the reinterview household roster could not be matched to the original household roster. These 'reject' cases were considered cases with discrepant results. Additional questions were asked to determine the reason for the discrepancy. Discrepant cases were then to be further classified as containing falsification, enumerator error, or respondent error.

Upon confirmation of discrepant results, the LCO management staff applied corrective action to the questionnaires and the enumerators, as appropriate.

During analysis, some cases were found to contain appropriate information to reject the case, yet no classification of the discrepancy is given. For this evaluation, errors were considered any discrepancies not attributed to respondent error. Error cases include data falsification, enumerator error, and unexplained cases.

## **1.3 How were data about Census 2000 reinterview selection and results obtained for the evaluation?**

Summary data files of the outlier enumerators from the administrative reinterview were delivered to the Decennial Statistical Studies Division (DSSD) in D908 files by the Technologies Management Office (TMO) on a weekly basis during the NFRU operation.

Summary data files of the cases selected for reinterview were delivered to DSSD in D191 files by TMO on a weekly basis during the NRFU operation. The D191 files are considered the master reference file for all cases selected for reinterview (D191 Master).

The D806 forms were sent to the National Processing Center (NPC) during and upon completion of the NRFU operation. Both key from image (KFI) and key from paper (KFP) technologies were used to capture the forms. Minimal data, including reinterview results and housing unit identification numbers, were captured from each form. Additional data were captured for a sample of forms. All data capture operations were subject to quality control measures.

The D191 forms are paper forms that were used to record batch level decision information. These forms were data captured at the NPC by KFI technology. Data from these forms are referred to as batch results.

## **2. METHODS**

### **2.1 How were forms identified for data capture?**

Every NRFU Reinterview form received at NPC went through the first step (Step 1) of the D806 data capture. Forms were identified by their housing unit identification number (Unit ID) and the preliminary reinterview decision (to accept or reject the case) was data captured for every form. Based on the results of Step 1 data capture and batch results, we selected a sample of cases for more complete data capture (Step 2).

For Step 2 data capture, some cases were considered certainty cases and others were selected by a 10 percent systematic sample of the forms (forms were not sorted in any particular order). There were three reasons a case might be selected with certainty to be in the sample of cases for Step 2:

- If the case was part of an administrative batch
- If there was any evidence of error in the batch
- If the preliminary reinterview decision on the form was “Reject.”

### **2.2 How were cases weighted to account for data capture?**

Results from Step 2 data capture were merged with files containing the complete Step 1 data capture results and all cases selected for Reinterview (D191 Master). This combined file was then priority unduplicated by Unit ID. (The unique identifier for Step 1 and Step 2 data capture was an image ID, as we may have had more than one form for a unique Unit ID.) We were not able to determine the sample type (random, administrative, supplemental) or the enumerator who completed the enumerator questionnaire in every case. In some cases the D806 was missing adequate information to make the linkage. In addition, some reinterview cases were received at the NPC that were not in the D191 Master file. This was a relatively small number of forms, and they were excluded from most of the analysis (see Table A in the Appendix).

Summaries for administrative case-level data are actual results. Since all administrative cases were selected with certainty for Step 2 data capture, results contain all administrative reinterview data available and are not sample-based results.

Case-level random reinterview and supplemental reinterview results are presented as estimates. Certainty cases were maintained with a weight of one. The remaining cases in the Step 2 data capture were weighted and combined with the certainty cases to produce data estimates.

The distribution of random, administrative, and supplemental cases selected for reinterview is 93.09 percent, 4.34 percent, and 2.57 percent, as identified by the D191 master file. When reporting a combined percentage for the reinterview operation, these weights are used to accurately represent the sample types in the relative magnitude they contribute to all the reinterview cases.

Care is taken to give percentages so administrative cases are not over-represented. Counts can be found in Table A in the Appendix.

## **2.3 How were enumerator summaries made?**

The data files used for this evaluation do not represent every enumerator. Since the data capture sample was based on forms, the 10 percent sample of forms mentioned above over-represents enumerators who had more cases selected for reinterview. The sample selection was not designed to make estimates of enumerator-level results. Summaries given are based on a 10 percent sample of forms and are provided for enumerators with fewer than ten forms selected for reinterview and those with ten or more. This division seems to be important for two reasons. The first reason has to do with the enumerators and their probability of having a case with complete data capture. With a ten percent sample, we anticipate most enumerators with ten or more cases in reinterview to be represented in the second category (ten or more). Because of the selection process for Step 2 data capture, not all enumerators have data represented in the first category of fewer than ten cases in reinterview. This category division also made sense in looking at the data. The more cases reinterviewed, the more likely an error was identified. This division helps show that impact and would help draw conclusions in the event that not all the case-level sample types were correctly assigned.

In order not to over-represent enumerators with certainty cases, including administrative cases and discrepant results, the file used for enumerator-level summaries is slightly different from the file used for case-level summaries. A file was created with a 10 percent sample of all forms, and estimates were not projected, but rather derived from available information, with all cases given an equal representation.

## **2.4 What analysis techniques were used?**

Using regression analysis, variables were tested using the dataset of a 10 percent sample of reinterview forms. The response variable that we used in testing the contribution of the proposed variables was an indicator that categorized each enumerator as either having a case in error or no cases in error. When D908, administrative outlier data, were not available for an enumerator, we assumed that the enumerator was included in the outlier tests, but not determined to be an outlier. Enumerator-level data, with certainty cases in the appropriate proportions, were used in the regression analysis.

We ran a stepwise discriminant regression analysis to look at the contribution of four D908 variables and five new variables. All two-way interactions were also included. The goal of the analysis was to determine the effect of the nine independent variables on our ability to detect errors in reinterview.

## **3. LIMITS**

### **3.1 Limitations to evaluation responses for all of the research questions.**

Reinterview was conducted in three of the Census 2000 enumeration operations: NRFU, List/Enumerate (L/E), and Update/Enumerate (U/E). There were incomplete data available from the L/E operation. The U/E operation was very small in terms of reinterview workload, compared to NRFU. Therefore it is unlikely that U/E reinterview data would significantly contribute to our analysis results. For these reasons and to ensure consistency, this evaluation used results from the NRFU reinterview only.

Originally, we thought that a look at the Accuracy and Coverage Evaluation (A.C.E.) Quality Assurance (QA) Check results would provide insight to additional variables that could be used in an administrative model. Because the reinterview program used different case-selection procedures (i.e., case selection was targeted), it was determined not appropriate to combine the results of the NRFU and A.C.E. Reinterview operations. Therefore, analysis of the variables used in the A.C.E. QA Check was excluded from this evaluation.

Average population per household was one of the characteristics tested in the administrative test and was used to identify outlier enumerators on the D908. Due to a processing problem, we were unable to identify when enumerators were outliers for average population in combination with one of the other four variables. Therefore, the influence of the average population flag is not included in this evaluation.

### **3.2 Data quality limitations**

From April 12, 2000 to April 16, 2000, the Operations Control System (OCS) system was at full capacity in loading NRFU data in the regional databases. We were warned that there may be some missed deliveries or inconsistent data in our deliveries during this time frame.

We noticed a few inconsistencies between the weekly D908 files we received and the paper D908 reports. The amount of inconsistent or missing data is unknown, but is assumed to be minimal.

During NRFU, there was no review of the D806 reinterview forms for completeness and accuracy either in the local census office, or at NPC prior to data capture. Missing data and inconsistencies on the forms contributed to there being a low percentage of cases (67.22) with complete reinterview results.

Summary information about the administrative batches and administrative cases was sometimes inconsistent with how the administrative reinterview should have been conducted. Over half of the administrative batches did not have any assigned administrative cases (of 35,591 enumerators with administrative batches, only 16,328 enumerators had administrative cases). We expected that all the administrative batches would contain administrative cases. We were unable to determine if the batch sample type variable or the case sample type variable was more accurate. For this evaluation, we assumed that the administrative cases were properly assigned and we only considered enumerators in administrative reinterview when an administrative case was selected. If the case sample type was not properly assigned, reinterview comparisons and conclusions would be less reliable.

### **3.3 Reinterview program limitations**

Once an enumerator was selected for administrative reinterview, the next ten enumerator questionnaires completed by that enumerator were selected as reinterview cases. There was no targeting for the specific cases. For example, if an enumerator was flagged for high deletes, there was no guarantee that a deleted case would be reinterviewed. For the administrative reinterview to achieve full effectiveness under this approach, we must assume, for example, that if an enumerator is an outlier for deleted cases, then that enumerator is likely to have errors in all the cases they complete (not just deleted cases). Review of Census 2000 NRFU data confirms that this assumption will not be valid for all, or even most, enumerators. Because there was no relation between the targeting characteristics and the selection of the reinterview cases, the results of our evaluation may not reflect the true relationship between work characteristics and the existence of discrepant results.

## 4. RESULTS

### 4.1 What was the Census 2000 NRFU Reinterview workload?

Data were provided to the DSSD from the Census 2000 OCS system maintained by TMO for the purpose of managing cases in the field. Each case assigned to reinterview was tracked through stages of completion. In NRFU, **2,546,359** cases were selected for reinterview. Each case can be attributed to being selected by random, administrative, or supplemental reinterview. A random batch generally consisted of seven reinterview cases and ten cases were to be selected for administrative and supplemental batches. If the random batch was not yet completed and the enumerator was selected to be in administrative reinterview, the whole batch was considered to be an administrative batch, even though individual cases within the batch might have been selected for both random and administrative reinterview.

The following table provides the selected NRFU reinterview workload and indicates the individual case sample type for each case.

Table 4.1.1 Totals and percent of cases selected for reinterview by sample type

|   | Total cases<br>selected for<br>reinterview | Random<br>(Sample Type 1) | Administrative<br>(Sample Type 2) | Supplemental<br>(Sample Type 3) |
|---|--|---------------------------|-----------------------------------|---------------------------------|
|   | 2,546,359                                  | 2,370,316                 | 110,597                           | 65,436                          |
| Percent of<br>cases selected<br>for reinterview | 100.00 %                                   | 93.09 %                   | 4.34 %                            | 2.57 %                          |

### 4.2 How many enumerators were selected for Reinterview?

The Decennial Management Division reported in its NRFU assessment report, **622,951** enumerators assigned as staff for NRFU. Reinterview data show that **465,769** enumerators were selected to have cases reinterviewed for some reason. The random reinterview was designed to reinterview a sample of cases for every enumerator. A case was selected randomly from the first ten cases completed. It is likely that the enumerators never selected for reinterview completed fewer than ten cases. Of enumerators who completed ten or more questionnaires (465,769 is a reasonable estimate), almost all enumerators had cases in random reinterview. Enumerators could be identified for administrative or supplemental reinterview at any time. Once cases were selected as administrative or supplemental reinterview, the enumerator was exempt from the random reinterview.

The following is a table of the enumerators selected for reinterview. The counts are of enumerators with one or more cases of the sample type specified in the column heading. The rows specify which enumerators are included in the category (all enumerators, those with less than ten cases selected for reinterview, and those with ten or more). The percentages represent the percent of the enumerators in that row that had any cases in the specified sample type. Enumerators may have reinterview cases of more than one sample type, so row percentages will not add to 100 percent.

Table 4.2.1 Enumerators selected for reinterview by sample type, categorized by total number of forms selected for reinterview for the enumerator

|   | Enumerators<br>selected for<br>reinterview | Random - 1 | Administrative - 2 | Supplemental - 3 |
|---|--|------------|--------------------|------------------|
| All Enumerators   | 465,769                                    | 465,513    | 16,328             | 8,829            |
| Row percent   |  | 99.95 %    | 3.51 %             | 1.90 %           |
| Column percent  | 100.00 %                                   | 100.00 %   | 100.00 %           | 100.00 %         |
| Enumerators with<br>fewer than 10 cases<br>in reinterview | 443,164                                    | 442,965    | 2,241              | 2,220            |
| Row percent   |  | 99.96 %    | 0.51 %             | .50 %            |
| Column percent  | 95.15 %                                    | 95.16 %    | 13.72 %            | 25.14 %          |
| Enumerators with<br>10 or more cases in<br>reinterview    | 22,605                                     | 22,548     | 14,087             | 6,609            |
| Row percent   |  | 99.75 %    | 62.32 %            | 29.24 %          |
| Column percent  | 4.85 %                                     | 4.84 %     | 86.28 %            | 74.86 %          |

More than 95 percent of the enumerators selected for reinterview had fewer than ten cases selected. It is consistent with the reinterview selection process that a higher percentage of enumerators with more than ten cases selected would have administrative or supplemental cases. Of the enumerators who had ten or more cases selected for reinterview, only 2,299 (10.17 percent) had no administrative or supplemental cases (see Table B, Reinterview case sample types by enumerator in the Appendix). Since the random reinterview was designed to select only seven cases, these results are not consistent with how random cases were to be selected. This is not a problem, except it shows another inconsistency of either the reinterview selection or recording. We continue to assume that case-level sample assignment is accurate and make sample type comparisons, although this may not be accurate in all cases.

Of the enumerators with an administrative case, 86.28 percent (14,087/16,328) had a total of ten or more cases selected. For enumerators with a supplemental case, 74.86 percent had ten or more cases selected. One reason for less than ten cases selected for administrative or supplemental reinterview would be that an enumerator did not complete ten or more enumerator questionnaires after selection for the reinterview. Employment with the Census Bureau may have been terminated or no additional work assigned.

#### **4.3 How many enumerators were identified as outliers by the administrative test?**

The test to determine the administrative sample was run 11 times during the NRFU operation, each week from May 7, 2000 to July 16, 2000. Each week an enumerator had completed questionnaires, their information was compared against the averages for their Crew Leader District (CLD). The D908 identified enumerators whose work was significantly different from the other enumerators in their CLD. The items compared were:

- Average population per household
- Delete Rate
- Population Count of One Rate
- Vacancy Rate
- Partial Interview Rate

The significance test for the average population per household was a two-tailed t-test. We looked for a high rate of occurrence for the rest of the variables and used a one-tailed Z-test of significance. Due to a processing problem, we were unable to identify when an enumerator was an outlier for average population in combination with one of the other four variables. We were able to determine when average population per household was the only variable that indicates the enumerator as an outlier, but with this limited information, comparisons would not be useful and are not included. The following table looks at the number of times each variable contributed to an enumerator being an outlier and the number of outlier enumerators for each of the variables.

Table 4.3.1 Number of times each of the D908 variables contributed to an enumerator being an outlier and the number of unique outlier enumerators for each variable

| Outlier Variable                  | Individual Flags<br>Identifying Outlier<br>Enumerators | Unique<br>Outlier Enumerators * |
|-----------------------------------|--|---------------------------------|
| High number of deleted units      | 296,479  | 109,211                         |
| High number of population of one  | 274,794  | 101,943                         |
| High number of vacant units       | 293,328  | 103,201                         |
| High number of partial interviews | 190,517  | 77,209                          |
| All variables                     | 1,055,118  | 291,441                         |

\* Enumerators could be outliers more than once for a variable and for more than one variable any given week.



There were **291,441** unique NRFU enumerators who were outliers for any reason during any of the 11 weeks. This is 62.57 percent of NRFU enumerators with completed work. The administrative model was designed to identify no more than 5 percent of enumerators as outliers for each variable, each week. The percentage of enumerators identified as outliers was higher than we anticipated.

Note: The L/E and U/E reinterview programs limited the administrative test to enumerators with 30 or more completed questionnaires. In NRFU, the enumerator could be an outlier for administrative reinterview based on any number of cases. If an enumerator completed work for more than one CLD, that enumerator could be an outlier in each CLD. Statistically, these situations make it more probable for an enumerator to be an outlier in NRFU.

#### **4.4 What percent of the enumerators identified as outliers by the administrative test were selected for administrative reinterview?**

The D908 outlier report was reviewed in the office by the OOS for Reinterview each week after it was printed. The OOS for Reinterview was instructed to place each enumerator in administrative reinterview unless there was a reasonable explanation for the characteristics being significantly different from the CLD average.

Of the **291,441** enumerators identified on the D908 reports, **14,696** had administrative cases selected for reinterview. This is **5.04 percent** of the outlier enumerators. The QA plan did not set an expectation of the percent of enumerators to be selected for administrative reinterview, however it was not anticipated the majority of the outlier enumerators would be justified and exempted from the administrative reinterview.

The low percentage of enumerators who had administrative cases selected makes it difficult to evaluate the effectiveness of the administrative variables. This evaluation continues to look at the effect of those variables, although this low percentage may make it difficult to draw conclusions. We recommend that in the future, the administrative reinterview component contain more specific expectations on the allowable amount of exceptions that can be made.

The way the reinterview process was set up, the OOS for Reinterview decision to put enumerators into administrative reinterview was not automated. The OOS for Reinterview, or their clerical staff, had to take the time to manually enter each enumerator into the system to select them for the administrative reinterview.

There were some administrative cases selected for reinterview for enumerators who weren't outliers on the D908. Two possible explanations are that either we were missing the D908 delivery or the cases were selected for administrative reinterview when they should have been categorized as supplemental.

Almost all (99.83 percent) of the outlier enumerators identified by the D908 reports had one or more cases selected to be in reinterview for any reason.

#### 4.5 What percent of the selected reinterview cases contain useable results?

The reinterview data provided above were obtained from the D191 files (master record of all reinterview cases). Not all cases selected for reinterview were completed. This was evidenced by blank forms, forms without original enumeration information, or forms with no reinterview data. Many comments indicated that cases were not completed because the original questionnaire was shipped from the LCO before the transcription was completed. Cases never received at NPC, forms missing the housing unit identification number, or forms containing inconsistent data also contributed to the number of cases without complete reinterview data.

Approximately **70 percent** of the reinterview forms received at NPC with any data had useable reinterview information. This is only **67.22 percent** of the total NRFU Reinterview workload.

Table 4.5.1 Percent of selected reinterview cases with complete reinterview data, by sample type

|  | Total   | Random Cases | Administrative Cases (actual) | Supplemental Cases |
|--|---------|--------------|-------------------------------|--------------------|
| Percent of selected reinterview cases with complete reinterview data | 67.22 % | 67.09 %      | 69.61 %                       | 50.58 %            |

The incompleteness rates are comparable for random and administrative sample types. Supplemental reinterview had a smaller proportion of cases with complete reinterview data. The percentage of supplemental cases missing original enumeration information was substantially higher compared to random and administrative reinterview. Sometimes reinterview was conducted by keeping the enumerator questionnaire with the D806 reinterview form and omitting the transcription. This seems most convenient with supplemental cases and may explain the lower completion rates. Enumerators are generally selected for supplemental reinterview when enumerator questionnaires have been recently turned into the LCO. The crew leader may be involved, or the reinterview staff may be alerted and the questionnaires may be hand-walked through the phone interview. Although a reinterview may have been completed, and results compared with the original enumeration, these cases are not maintained in the analysis, since there is no documented comparison.

#### 4.6 What percent of the completed reinterview cases contained discrepant results?

A case was to be determined to contain discrepant results when the housing unit status from the reinterview did not match the original status and/or at least 50 percent of the reinterview household roster did not match to the original household roster. The reinterview staff was instructed to “reject” those cases. Cases incorrectly “accepted” are categorized here as unexplained.

If the reinterviewer determined that the original enumerator deliberately entered falsified data, then the case was identified as falsification. The case was identified as enumerator error if it was determined that the incorrect data were mistakes made by the enumerator. The reinterview cases were identified as respondent error if it was determined that the incorrect data were mistakes made by the respondent. Some cases with discrepancies were not identified as falsification, enumerator error, or respondent error and are categorized as unexplained. For this evaluation we are concerned about discrepancies caused by enumerator error, both intentional and unintentional. Errors are considered any discrepancy not attributed to respondent error. An estimated **9.51 percent** of the total NRFU reinterview cases contained errors.

Table 4.6.1 Error rates for completed reinterview cases by sample type and type of error

| Categorization of discrepant results            | Random Cases | Administrative Cases<br>(Actual) | Supplemental Cases |
|---|--------------|----------------------------------|--------------------|
| Enumerator Error                                | 4.80 %       | 4.83 %                           | 5.61 %             |
| Falsification                                   | 1.72 %       | 1.81 %                           | 2.39 %             |
| Unexplained                                     | 2.86 %       | 2.98 %                           | 3.25 %             |
| Respondent error                                | 1.73 %       | 1.92 %                           | 2.27 %             |
| Enumerator error, falsification, or unexplained | 9.42 %       | 9.67 %                           | 11.30 %            |

Practically, random reinterview was just as effective as administrative reinterview in identifying cases with error. Supplemental reinterview identified more error cases than administrative or random reinterview.

We have no reason to suspect that the incidence of error would have been any greater or less for cases that were selected for reinterview, but did not contain complete reinterview results.

#### **4.7 What percent of enumerators with complete reinterview results had any cases with errors?**

Since data for only a sample of forms were data captured, we did not obtain data for every enumerator. The data capture sample was based on forms and not intended to be used to estimate enumerator-level statistics. We continue to give enumerator-level statistics, because the administrative test was performed at the enumerator level. The enumerator information that follows is not an estimate of all enumerators represented in reinterview, however it can provide useful insight to enumerator summaries based on the sample of forms with complete reinterview results. Enumerators with less cases in reinterview are under-represented in these summaries, but

the summaries reflect enumerators in the magnitude that they contributed to completing NRFU work.

Approximately 95 percent of enumerators had fewer than ten cases selected for reinterview. Of the enumerators in sample with fewer than ten cases selected for reinterview, **10.76 percent** had one or more cases with errors, excluding respondent error. Of the enumerators in sample with ten or more cases selected for reinterview, **13.54 percent** of enumerators had one or more cases with errors. Enumerators may have had cases attributed to more than one of the categories below.

#### 4.7.1 Percentage of enumerators with completed reinterview cases by categories of errors

| Categorization of errors | Percent of<br>enumerators with<br>fewer than<br>10 cases selected for<br>reinterview | Percent of<br>enumerators with<br>10 or more cases<br>selected for<br>reinterview |
|--------------------------|--|---|
| Enumerator Error         | 5.60 %   | 6.78 %  |
| Falsification            | 1.95 %   | 2.98 %  |
| Unexplained              | 3.34 %   | 4.20 %  |
| Total                    | 10.76 % *  | 13.54 % *   |

\* These percentages do not sum to 10.76 percent and 13.54 percent, as enumerators may have had cases in more than one of the error categories.

Consistent with the case-level summaries in section 4.6, most of the errors were categorized as enumerator error. The more cases reinterviewed, the more likely to identify cases containing errors.

#### 4.8 How effective was the administrative reinterview in identifying enumerators with errors?

Based on the sample of forms, the following types of cases were influential in identifying enumerators with cases in error. Row percentages may add up to more than 100 percent since more than one case may have been found in error for an enumerator.

Table 4.8.1 Percent of cases contributing to enumerators with error cases, by sample type and number of forms selected for reinterview.

|  | Random Case | Administrative Case | Supplemental Case |
|--|-------------|---------------------|-------------------|
| Fewer than 10 cases selected for reinterview | 99.42 %     | 0.32 %              | 0.26 %            |
| 10 or more cases selected for reinterview    | 45.83 %     | 39.89 %             | 15.96 %           |

For enumerators with fewer than ten cases selected for reinterview, almost all of the error cases were identified by a random reinterview case. For enumerators with ten or more cases selected for reinterview, the percent of cases in error identified by the administrative reinterview jumps from 0.32 percent to 39.89 percent. These results are expected because, as reported earlier, very few of the enumerators in the ‘fewer than 10 cases’ category had administrative cases selected and most of the enumerators in the ‘10 or more cases’ category had administrative cases selected.

In section 4.6, we indicated that at a reinterview *case* level, administrative reinterview is no more effective than random reinterview. The following table follows the progression of enumerators with administrative cases, or administrative enumerators.

Table 4.8.2 Frequencies of administrative enumerators across steps of reinterview and data capture process

|                                  | Administrative Enumerators |
|----------------------------------|----------------------------|
| Selected for reinterview         | 16,328                     |
| Any Step 2 data capture results  | 15,360                     |
| Any complete reinterview results | 15,020                     |
| Any cases in error               | 7,824                      |

Results at an enumerator level yield different results than the case-level summaries. We found errors in **52.09 percent** of administrative enumerators with complete reinterview data. Administrative case-level error rates were 9.67 percent.

The numbers above include all available data captured for these enumerators. Data capture was designed to obtain all administrative cases and any cases with any apparent evidence of discrepancies. Cases in administrative batches are over-represented, so no direct comparison can be made between the administrative error rates and random reinterview rates for these specific enumerators. Error rates for enumerators with random reinterview will be close to the overall enumerator rates given in the previous section.

#### **4.9 What is the contribution of each of the Census 2000 variables in identifying cases with discrepant results?**

Table 4.9.1 Percentages of reinterviewed enumerators identified as outliers, by variable in the administrative test and with cases in error (enumerator level dataset)

| Outlier Variable                  | Percent of NRFU identified as outliers | Percent of Enumerators with Error Cases |
|-----------------------------------|--|---|
| High number of deleted units      | 29.83                                  | 33.03                                   |
| High number of population of one  | 28.76                                  | 29.00                                   |
| High number of vacant units       | 28.83                                  | 28.53                                   |
| High number of partial interviews | 21.59                                  | 21.64                                   |

The delete variable was the best indicator of when an enumerator was in error. Of the enumerators with error cases, 33.03 percent were identified as outliers by the high delete variable.

The percentages above are representative of all the forms. The rates differ from section 4.3.1 because they under-represent enumerators with fewer reinterview cases completed. We also reviewed enumerators with ten or more cases selected for reinterview and the enumerators in administrative reinterview. The proportion of outlier enumerators increases (high deletes, population of one, and vacants are close to 40 percent), however error rates are similar to overall rates of enumerators in error.

In order to determine the effect of the variable, and not just the effect of the number of forms, we performed regression analysis. Results are in section 4.11.

#### **4.10 How do training and personal characteristics of the enumerators with discrepancies compare to all NRFU enumerators?**

We obtained information from the Preappointment Administrative Management System and the Automated Decennial Administrative Management System (PAMS/ADAMS) database about hours in training, education level, test score, any foreign language experience, and prior enumeration experience. We reviewed the data and based on the distribution of all enumerators created categories relatively similar in size, as follows:

Table 4.10.1 New variable descriptions and categories

|                              |  |
|------------------------------|--|
| Hours in training            | Less than 22 hours, 22 - 26 hours, and more than 26 hours          |
| Education level              | High school, some education past high school, and college graduate |
| Test score                   | Less than 90, 90 and above   |
| Additional language ability  | Yes/No   |
| Prior enumeration experience | Yes/No   |

When the frequencies were compared between all NRFU enumerators in the sample and enumerators with discrepancies, proportions were close in every category.

Table 4.10.2 Distribution of all NRFU enumerators and enumerators with errors for each new variable

|                                       | Percent of all NRFU<br>enumerators with<br>reinterview results | Percent of<br>enumerators with<br>error cases |
|---------------------------------------|--|---|
| Less than 22 hours in training        | 37.95  | 37.05   |
| 22 - 26 hours in training             | 47.09  | 47.48   |
| More than 26 hours in training        | 14.96  | 15.47   |
| High school                           | 50.85  | 52.42   |
| Some education past high school       | 18.63  | 18.95   |
| College graduate                      | 30.53  | 28.63   |
| Test score less than 90               | 38.24  | 42.18   |
| Test score 90 and above               | 61.76  | 57.82   |
| Additional language ability (Yes)     | 69.62  | 69.06   |
| No additional language ability        | 30.38  | 30.94   |
| Prior enumeration experience<br>(Yes) | 81.88  | 81.29   |
| No prior enumeration experience       | 18.12  | 18.71   |

These variables do not have a practical impact on whether or not an enumerator had error cases, but we performed regression analysis to look for significant interactions.

#### **4.11 Could any of the new variables be used to predict or target enumerators who would be more likely to have discrepant results?**

A stepwise discriminant regression model was run using nine independent variables to predict enumerators who had any discrepant cases. The five new variables from PAMS/ADAMS were analyzed along with four indicators of the D908 characteristics. The D908 indicator variables indicated if the enumerator was ever an outlier for that characteristic.

In order to create a good prediction model, the R-squared value should be fairly close to 1. Because of the size of our data files, all of the variables were statistically significant. However, the R-squared value was less than 0.01. A low R-squared value indicates that very little of the difference in error rates can be attributed to the predictor variables. See the regression output in Table E in the Appendix.

The correlation is very low and although these variables may be slightly more effective than taking a purely random reinterview, many enumerators without error results would be identified as outliers. For instance, if nationwide we had targeted 100 enumerators because they had a high delete rate, our selection of more of their cases to be reinterviewed might well find that 97 of them were no longer working in high delete rate areas but three of them were curbstoning by misclassifying units as deletes. In this case the administrative reinterview would be very effective in finding curbstoners but not necessarily show that delete rate was a significant factor in identifying curbstoning.

## **5. CONCLUSIONS AND RECOMMENDATIONS**

Targeting enumerators using the variables researched (high number of deleted units, high number of population of one, high number of vacant units, high number of partial interviews, hours in training, education level, test score, additional language ability, and prior enumeration experience) does not result in a higher detection rate. Results also indicate that 9.51 percent of the cases reinterviewed contained errors. The more cases reinterviewed for an enumerator, the higher the instance of identifying (and correcting) discrepant results.

Other studies and academic experts<sup>1</sup> have shown that targeted reinterview can be very beneficial. Targeting case selection in administrative reinterview may be a more effective way to use the variables to identify enumerators with cases in error.

Additional research would be beneficial in several areas.

---

<sup>1</sup> See Reference #8



Further analysis of the D806 results could determine a classification of errors, such as an occupied unit classified as a delete, an incorrect roster, etc. These classifications could then be compared to the outlier variables and analyzed for correlations.

The D908 data were summarized into one indicator flag. Additional data are available to do analysis on cutoff scores, weekly effects, the number of short and long forms completed, and combined effect of the administrative variables. This would help in understanding the distribution and validating assumptions made in the design of the administrative model. We could look at the impact of enforcing a minimum of 30 completed cases or increasing the levels for outlier cutoffs, so that less enumerators are identified as outliers. The percent of outlier enumerators must be reduced to have an effective administrative reinterview.

## REFERENCES

- 1) "Quality Assurance Software Specification for Generating the Random Sample for Detection of Fabrication and other Performance Errors during the Census 2000 Nonresponse Followup Operation." Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series II-5. May 26, 1999.
- 2) "Quality Assurance Software Specifications for Performing the Administrative Test for Detection of Fabrication or other Performance Errors during the Census 2000 Nonresponse Followup Operation." Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series II-8. June 6, 1999.
- 3) "Revised: Quality Assurance Specifications for the Census 2000 Nonresponse Followup Operation." Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series II-9R. August 28, 1999.
- 4) Rosemary Byrne, "Census 2000 A.C.E. QA Targeting Reports Specifications for Person Interviews." Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series Chapter S-QA-06, Feb 7, 2001.
- 5) "Data Requirements for Reinterview Documentation for the Operations of Nonresponse Followup, List/Enumerate, and Rural Update Enumerate." Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series II-18. January 4, 2000
- 6) "Request for Data Capture and Cost Estimate of Step 1 D-806, Reconciliation and Reinterview Questionnaires (wand and key from paper)." Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series II-M2-3. August 13, 2001.

7) “Request for Imaging and Data Capture of Reinterview and Reconciliation Questionnaire, D-806.” Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series II-M2-4. November 27, 2001.

8) “Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection and Repair of Its Effects.” AAPOR Standard Committee. Draft 3, April 21, 2003. <http://www.aapor.org/interviewfalse.pdf>

## Appendix

|  |      |
|--|------|
| Form D-908 (NRFU) Administrative Reinterview Trouble Report .....                          | A-2  |
| Form D-191 (NRFU-RI) Reinterview Control Record .....                                      | A-3  |
| Form D-806 Reinterview and Reconciliation Questionnaire .....                              | A-4  |
| Table A Unduplicated case level counts through the reinterview process by case sample type | A-8  |
| Table B Reinterview case sample types by enumerator .....                                  | A-12 |
| Table C Results from discriminant regression analysis .....                                | A-13 |

THIS REPORT CONTAINS INFORMATION, THE RELEASE OF WHICH IS PROHIBITED BY THE PRIVACY ACT OF 1974, AND IS FOR THE BUREAU OF THE CENSUS OFFICIAL USE ONLY.

**D-908 (NRFU)**  
(08/99)

U.S. DEPARTMENT OF COMMERCE  
BUREAU OF THE CENSUS

**ADMINISTRATIVE REINTERVIEW TROUBLE REPORT**  
OPERATION : NONRESPONSE FOLLOWUP  
U.S. CENSUS 2000

RCC/LCO: 2599/2515

CONNECTICUT/HARTFORD CT

FOS: 01

CLD: 0101

Create Date/Time: 06/15/1999 14:00

Print Date/Time: 02/25/2000 08:13

Page 1 of 1

| Enumerator/<br>SSN<br>(1) | Last 7 Days            |                        |                   |                          |                      | Form Type                      |                                 | Action  |   |
|---------------------------|------------------------|------------------------|-------------------|--------------------------|----------------------|--------------------------------|---------------------------------|---|---|
|                           | High<br>Vacants<br>(2) | High<br>Deletes<br>(3) | High<br>PI<br>(4) | Avg Pop/<br>ER HI<br>(5) | High<br>POP 1<br>(6) | Number of<br>long forms<br>(7) | Number of<br>short forms<br>(8) | Reinterview<br>1) Yes 2) No<br>(9)                      | Justification<br>(options below)<br>(10)  |
|                           |                        |                        | 30.0%             | 4.5                      |                      | 5                              | 15                              | 1. <input type="checkbox"/> 2. <input type="checkbox"/> | 1. <input type="checkbox"/> 2. <input type="checkbox"/> 3. <input type="checkbox"/> |
|                           |                        | 31.3%                  |                   |                          |                      | 4                              | 12                              | 1. <input type="checkbox"/> 2. <input type="checkbox"/> | 1. <input type="checkbox"/> 2. <input type="checkbox"/> 3. <input type="checkbox"/> |
| <b>CLD AVERAGE</b>        | 12.5%                  | 8.8%                   | 15.0%             | 2.6                      | 10.0%                | 0                              | 18                              |   |   |

**Justification:**

1. Enumerator no longer working in CLD.
2. Characteristic consistent with area (for example, the area has a large number of seasonal vacants).
3. Other (specify) \_\_\_\_\_

A-2

D-191 (NRFU-RI)  
(02/00)

U.S. DEPARTMENT OF COMMERCE  
BUREAU OF THE CENSUS

REINTERVIEW CONTROL RECORD  
OPERATION: NRFU REINTERVIEW  
U.S. CENSUS 2000

RCC: 2199/BOSTON

LCO: 2145/Utica

Enumerator Name:

Print Date/Time: 05/12/2000 14:56

Page 1 of 1

| Reinterview<br>Number<br><br>(1) | CLD<br><br>(2) | ID<br>Number<br><br>(3) | Preliminary<br>Decision   | Final<br>Decision   |  |  |
|----------------------------------|----------------|-------------------------|---|---|--|--|
|                                  |                |                         | Evidence of Data<br>Falsification?<br>1. Yes<br>2. No, Accept/Noninterview<br>3. No, Enumerator Error<br>4. No, Respondent Error<br><br>(4) | Evidence of<br>Data<br>Falsification?<br>1. Yes<br>2. No<br><br>(5)       | Evidence of<br>Performance<br>Error?<br>1. Yes<br>2. No<br><br>(6) | FOS/AMFO Action Taken<br>1. Enumerator released<br>2. Enumerator resigned<br>3. Enumerator warned/advised<br>4. Additional cases selected<br>5. No action taken<br><br>(7) |
| a.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 | 1 <input type="checkbox"/> 2 <input type="checkbox"/>                     | 1 <input type="checkbox"/> 2 <input type="checkbox"/>              | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/>                                     |
| b.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 | Remarks:<br><br>_____<br><br>_____<br><br>_____<br><br>_____<br><br>_____ |  |  |
| c.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |
| d.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |
| e.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |
| f.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |
| g.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |
| h.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |
| i.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |
| j.                               |                |                         | 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>                                 |   |  |  |

OOS Signature:

Date:



**REINTERVIEW AND  
RECONCILIATION  
QUESTIONNAIRE**  
United States Census 2000

Notice - Response to this inquiry is required by law (title 13, U.S. Code). By the same law, your report to the Census Bureau is **confidential**. It may be seen only by sworn Census employees and may be used only for statistical purposes.

**Section 1 - IDENTIFICATION**

|        |            |             |
|--------|------------|-------------|
| 1. LCO | 2. Tract   | 3. Block    |
| 4. AA  | 5. Unit ID | 6. Map spot |

**7. MAILING ADDRESS/EMERGENCY ADDRESS/LOCATION DESCRIPTION**

(House number, street number, apartment number, or rural route, box number)

City State ZIP Code

**Section 2 - ORIGINAL ENUMERATOR  
INFORMATION****8. METHOD OF COMPLETION**1 ☐ Personal visit 2 ☐ Telephone**9a. Original respondent's name**

|              |           |        |  |
|--------------|-----------|--------|--|
| b. Telephone | Area code | Number | <input type="checkbox"/> No phone number |
|--------------|-----------|--------|--|

**c. Best time to call**1 ☐ Day  
2 ☐ Evening  
3 ☐ Either**d. Respondent —**1 ☐ lived here on April 1, 2000  
2 ☐ moved in after April 1, 2000  
3 ☐ is a neighbor or other**10. Enumerator's name**

Notes

**Section 3 - TELEPHONE INTRODUCTION -  
READ BOLD ONLY***(See Job Aid for Personal Visit Introduction)***A. Hello, my name is (Your name) and I'm calling for the United States Census Bureau. May I speak to (Original respondent in section 2, item 9a)?**

- If the original household respondent (9d is marked 1 or 2) is not available, ask to speak to a household member who is at least 15 years old. Then, read the Privacy Act Notice.
- If the original proxy (9d is marked 3) is not available, ask what time would be best to call back. Do not continue with B2 until/unless you reach the original proxy.

If item 9d is marked 1 or 2 - **READ:****B1. Have I reached (Read address in section 1, item 7)? If NO - Excuse me. I might have dialed the wrong number. Is this (Read area code and phone number in section 2, item 9b)?**

- ☐ Respondent phone number and address do not agree - Ask if respondent has ever lived at the address in section 1, item 7. If NO, end phone interview and send to Personal Reinterview. If YES, go to item C.

If item 9d is marked 3 - **READ:****B2. I'm calling in reference to information provided for (Read address in section 1, item 7).****C. We are checking the accuracy of the Census, and I'd like to ask you a few questions. This interview should take about 3 minutes. Complete items 11a, 11b, 12, and 13. Then, skip to item 16.****Section 4 - REINTERVIEW INFORMATION****11a. Reinterview respondent's name**

|              |           |        |  |
|--------------|-----------|--------|--|
| b. Telephone | Area code | Number | <input type="checkbox"/> No phone number |
|--------------|-----------|--------|--|

**12. Who is the reinterview respondent?  
Mark (X) ONE box**

- 1 ☐ Original household respondent
- 2 ☐ Different household respondent
- 3 ☐ Original proxy
- 4 ☐ Other

**13. TYPE OF REINTERVIEW - 6 telephone attempts must be made before any personal visits.**

| Telephone |     |      |      | Personal visit |     |      |      |
|-----------|-----|------|------|----------------|-----|------|------|
| Date      |     |      | Time | Date           |     |      | Time |
| Month     | Day | Year |      | Month          | Day | Year |      |
|           |     |      | a.m. |                |     |      | a.m. |
|           |     |      | p.m. |                |     |      | p.m. |
|           |     |      | a.m. |                |     |      | a.m. |
|           |     |      | p.m. |                |     |      | p.m. |
|           |     |      | a.m. |                |     |      | a.m. |
|           |     |      | p.m. |                |     |      | p.m. |
|           |     |      | a.m. |                |     |      | a.m. |
|           |     |      | p.m. |                |     |      | p.m. |
|           |     |      | a.m. |                |     |      | a.m. |
|           |     |      | p.m. |                |     |      | p.m. |

If unable to contact a knowledgeable person after 6 telephone and 3 personal visit attempts, skip to item 23 on page 4.

| Part A - ORIGINAL ENUMERATOR RESPONSE  |  |  |  | Part B - REINTERVIEW RESPONSE  |  | Part C - MATCH RESULTS |  |
|--|--|--|--|--|--|------------------------|--|
| <b>14. UNIT STATUS:</b><br>1 <input type="checkbox"/> Occupied<br>2 <input type="checkbox"/> Vacant<br>3 <input type="checkbox"/> Delete/Not a living quarters   |  | <b>16. On April 1, 2000, was this unit -</b><br>1 <input type="checkbox"/> Occupied? Ask 17<br>2 <input type="checkbox"/> Vacant? Skip to 20<br>3 <input type="checkbox"/> Not a living quarters? Skip to 20 |  | <b>20. UNIT STATUS (Mark X)</b><br>Match      Nonmatch   |  |                        |  |
| <b>15. ORIGINAL HOUSEHOLD ROSTER - Refer to question 1 of the original Form D-1(E) or 2(E) Enumerator Questionnaire (Use D-1(E) or D-2(E) continuation forms as necessary). Continue on page 3 if needed.</b><br>(a) |  | <b>17. Did you or anyone in this household, live here on Saturday, April 1, 2000?</b><br>Ask only if 9d is marked 1 or 2<br>1 <input type="checkbox"/> Yes<br>2 <input type="checkbox"/> No                  |  | <b>21. HOUSEHOLD ROSTER</b><br>Match (Enter person numbers from item 15a)      Nonmatch (Mark X) |  |                        |  |
| <b>PERSON 1</b> First name      Middle initial<br>Last name  |  | <b>PERSON 1</b> First name      Middle initial<br>Last name  |  | Match      Nonmatch  |  |                        |  |
| <b>PERSON 2</b> First name      Middle initial<br>Last name  |  | <b>PERSON 2</b> First name      Middle initial<br>Last name  |  | Match      Nonmatch  |  |                        |  |
| <b>PERSON 3</b> First name      Middle initial<br>Last name  |  | <b>PERSON 3</b> First name      Middle initial<br>Last name  |  | Match      Nonmatch  |  |                        |  |
| <b>PERSON 4</b> First name      Middle initial<br>Last name  |  | <b>PERSON 4</b> First name      Middle initial<br>Last name  |  | Match      Nonmatch  |  |                        |  |
| <b>PERSON 5</b> First name      Middle initial<br>Last name  |  | <b>PERSON 5</b> First name      Middle initial<br>Last name  |  | Match      Nonmatch  |  |                        |  |
| <b>PERSON 6</b> First name      Middle initial<br>Last name  |  | <b>PERSON 6</b> First name      Middle initial<br>Last name  |  | Match      Nonmatch  |  |                        |  |
| <b>PERSON 7</b> First name      Middle initial<br>Last name  |  | <b>PERSON 7</b> First name      Middle initial<br>Last name  |  | Match      Nonmatch  |  |                        |  |
|  |  | <b>19. TOTAL REINTERVIEW ROSTER SIZE - Add in any extra names to item 18 from the reinterview roster on page 3 →</b>   |  | <b>22. TOTAL NON-MATCHES →</b>   |  |                        |  |



| Section 5 - REINTERVIEW AND RECONCILIATION - Continued |                |   |                |  |                   |
|--|----------------|---|----------------|--|-------------------|
| Part A - ORIGINAL ENUMERATOR RESPONSE                  |                | Part B - REINTERVIEW RESPONSE   |                | Part C - MATCH RESULTS   |                   |
| 15. ORIGINAL HOUSEHOLD ROSTER - Continued              |                | 18. REINTERVIEW HOUSEHOLD ROSTER - Continued  |                | 21. HOUSEHOLD ROSTER - Continued                               |                   |
| (a)  |                | (b)   |                | Match (Enter person numbers from item 15a)                     | Nonmatch (Mark X) |
| PERSON 8 First name                                    | Middle initial | PERSON 8 First name   | Middle initial |  |                   |
| Last name  |                | Last name   |                |  |                   |
| PERSON 9 First name                                    | Middle initial | PERSON 9 First name   | Middle initial |  |                   |
| Last name  |                | Last name   |                |  |                   |
| PERSON 10 First name                                   | Middle initial | PERSON 10 First name  | Middle initial |  |                   |
| Last name  |                | Last name   |                |  |                   |
| PERSON 11 First name                                   | Middle initial | PERSON 11 First name  | Middle initial |  |                   |
| Last name  |                | Last name   |                |  |                   |
| PERSON 12 First name                                   | Middle initial | PERSON 12 First name  | Middle initial |  |                   |
| Last name  |                | Last name   |                |  |                   |
| PERSON 13 First name                                   | Middle initial | PERSON 13 First name  | Middle initial |  |                   |
| Last name  |                | Last name   |                |  |                   |
| PERSON 14 First name                                   | Middle initial | PERSON 14 First name  | Middle initial |  |                   |
| Last name  |                | Last name   |                |  |                   |
|  |                | TOTAL - Reinterview roster size for this page only - Include this number in page 2, item 19 → |                | TOTAL this page only - Include this total in page 2, item 22 → |                   |
| Remarks  |                |   |                |  |                   |
|  |                |   |                |  |                   |
|  |                |   |                |  |                   |
|  |                |   |                |  |                   |
|  |                |   |                |  |                   |
|  |                |   |                |  |                   |



## Section 5 - REINTERVIEW AND RECONCILIATION - CONTINUE

### Part D - DECISION

#### 23. PRELIMINARY DECISION - Mark (X) one box

- 1 ☐ Accept - Mark "Accept" if housing status in item 20 is a match AND 50% or more of the roster in column 21 is a match. Thank the respondent and conclude the interview. Then skip to item 26.
- 2 ☐ Reject - Mark "Reject" if housing status in item 20 is a nonmatch OR less than 50% of the roster in column 21 is a match. Go to item 24.
- 3 ☐ Noninterview - Mark "Noninterview" if unable to contact a knowledgeable respondent after 6 phone and 3 field attempts. End interview. Skip to item 26.

### Part E - ACCOUNTABILITY

#### 24. To the best of your knowledge, has anyone from the U.S. Census Bureau recently contacted your household either by personal visit or telephone and conducted an interview?

- 1 ☐ Yes - Ask 25
- 2 ☐ No - Thank the respondent and conclude the interview. Then skip to item 26.

#### 25. Explanation - DO NOT ask this question unless item 12, box 1, 2, or 3 is marked. The original questionnaire showed (Read the original enumerator response - item 14 or 15, or both) and I have (Read the reinterview response - item 16 or 18, or both), can you think of any explanation of why there is a difference? Continue in Remarks section below, if needed.

---

---

---

---

Thank the respondent and conclude the interview. Complete items 26-29.

### Part F - CONCLUSION

#### 26. Does evidence indicate data falsifications?

- 1 ☐ Yes ☐ No - Mark (X) ONE box.
- 2 ☐ Accept/Noninterview
- 3 ☐ Enumerator Error
- 4 ☐ Respondent Error

#### 27. REINTERVIEW STATUS

- 1 ☐ Complete Noninterview
- 2 ☐ Refusal
- 3 ☐ Unable to locate
- 4 ☐ Other - Explain in "Remarks" section below.

28. Reinterviewer's name (Print) \_\_\_\_\_

#### 29. CERTIFICATION

a. Reinterviewer's signature

b. Date

I certify that the entries I have made on this questionnaire are true and correct to the best of my knowledge.

Remarks

---

---

---

---

---

---

**Table A Unduplicated case level counts through the reinterview process by case sample type**

| Ref | Description   | Unique Unit IDs                         | Reinterview Case Sample Type            |                                    |                              |                      | Source   |
|-----|---|---|---|------------------------------------|------------------------------|----------------------|--|
|     |   |   | Random                                  | Admin                              | Supp                         | Unknown              |  |
| 1   | Cases in NRFU Reinterview   | 2,507,836                               |   |                                    |                              |                      | DMD Operational Summary  |
| 2   | Cases in NRFU Reinterview   | 2,546,344                               | 2,204,041                               | 276,832                            | 65,436                       |                      | DMD Census 2000 Cost and Progress Report   |
| 3   | Cases in NRFU Reinterview -<br>Samp_typ<br>Bat_samp   | 2,546,359<br>2,546,359                  | 2,370,316<br>2,204,065                  | 110,597<br>276,832                 | 65,436<br>65,436             |                      | 191 Files from TMO (Note: this may understated, as a few deliveries were lost)<br>d191.master_b<br>where op='NRFU' |
| 4   | D806 forms from Step 1 reviewed for sample selection  | 2,419,540                               | 2,239,479                               | 102,768                            | 60,889                       | 16,404               | comb.undid<br>v8061='Yes'<br>freq samp_typ   |
| 5   | Certainty cases selected for Step 2 data capture<br>- falsification<br>- prelimdec (reject)<br>- error in 191 batch<br>- Admin batch (and not selected by 10% or other above) | 50,495<br>174,245<br>172,348<br>177,548 | 44,012<br>156,039<br>155,557<br>106,855 | 4,328<br>7,625<br>12,194<br>70,687 | 2,152<br>3,918<br>4,589<br>0 | 3<br>6,663<br>8<br>6 | comb.undid<br>v806S='Yes'<br>freq samp_typ*enumsamp  |

**Table A Unduplicated case level counts through the reinterview process by case sample type (continued)**

| Ref | Description  | Unique Unit IDs | Reinterview Case Sample Type |          |        |         | Source  |
|-----|--|-----------------|------------------------------|----------|--------|---------|---|
|     |  |                 | Random                       | Admin    | Supp   | Unknown |   |
| 6   | Cases selected for Step 2 data capture                 | 786,552         | 652,735                      | 102,761  | 15,816 | 15,240* | comb.undid<br>v806S='Yes'<br>freq samp_typ<br>* note sample selection was<br>done without case<br>unduplication |
| 7   | Certain = 10%  |                 | 190,272                      | (7,927)  | 5,157  | (8,560) |   |
| 8   | Certain = Yes  |                 | 462,463                      | (94,834) | 10,659 | (6,680) |   |
| 9   | Cases captured in step 2                               | 790,610         | 652,300                      | 102,717  | 15,815 | 19,778  | comb.undid<br>v8062='Yes'<br>freq samp_typ  |
| 10  | Certain = 10%  |                 | 190,127                      | 7,922    | 5,157  |         | comb.undid<br>v8062='Yes'<br>freq samp_typ*certain  |
| 11  | Certain = Yes  |                 | 462,173                      | 94,795   | 10,658 |         |   |
| 12  | Step 2 cases determined to have a complete reinterview | 570,842         | 472,798                      | 76,985   | 9,780  | 11,279  | comb.undid<br>cljrd='4=Reinterview'<br>freq samp_typ  |
| 13  | Certain = 10%  |                 | 133,981                      |          | 2,650  |         |   |
| 14  | Certain = Yes  |                 | 338,817                      |          | 7,130  |         |   |
| 15  | Estimated (total) cases to have complete reinterview   | 1,711,563       | 1,590,199                    | 76,985   | 33,100 | 11,279  | rand - multiplier 9.34<br>supp - multiplier 9.80  |



**Table A Unduplicated case level counts through the reinterview process by case sample type (continued)**

| Ref | Description                                | Unique Unit IDs | Reinterview Case Sample Type |                  |         |                  | Source   |
|-----|--|-----------------|------------------------------|------------------|---------|------------------|--|
|     |  |                 | Random                       | Admin            | Supp    | Unknown          |  |
| 16  | % of Step 2 cases with reinterview results | 70.62 %         | 71.01 %                      | 74.95 % (actual) | 54.36 % | 57.03 % (actual) | Admin is line 15/line 9<br>Rand/Supp is line 15/<br>((multiplier*line10)+<br>line 11)  |
| 15  | % of cases selected for reinterview        | 67.22 %         | 67.09 %                      | 69.61 %          | 50.58 % |                  | Line 15/line 3<br>Note: These percentages do not count for forms intentionally not processed at NPC and may slightly under-represent actual percentages (differences less than .15%) |
| 16  | Step 2 cases containing discrepant results | 150,774         | 133,502                      | 8,882            | 3,408   | 4,982            | comb.undid<br>error in ('Yes', 'YRe')<br>freq samp_typ   |
| 17  | Respondent Error                           | 24,431          | 21,533                       | 1,478            | 603     | 817              | comb.undid<br>errreason*samp_typ   |
| 18  | Enumerator Error                           | 74,103          | 66,260                       | 3,719            | 1,599   | 2,525            |  |
| 19  | Falsification                              | 26,757          | 23,754                       | 1,391            | 709     | 903              |  |
| 20  | Unexplained                                | 25,483          | 21,955                       | 2,294            | 497     | 737              |  |

**Table A Unduplicated case level counts through the reinterview process by case sample type (continued)**

| Ref | Description                   | Unique Unit IDs | Reinterview Case Sample Type |         |       |         | Source                                       |
|-----|-------------------------------|-----------------|------------------------------|---------|-------|---------|--|
|     |                               |                 | Random                       | Admin   | Supp  | Unknown |  |
| 21  | Step 2 cases containing error | 126,791         | 112,351                      | 7,448   | 2,815 | 4,177   | comb.undid<br>error = 'Yes'<br>freq samp_typ |
| 22  | Certain = 10%                 |                 | 5,087                        | (174)   | 108   | (274)   |  |
| 23  | Certain = Yes                 |                 | 107,264                      | (7,274) | 2,707 | (3,903) |  |

**Table B Reinterview sample types by enumerator**

The Case variable is a 3 position variable, indicating any random cases, administrative cases, and supplemental cases, in that order. There were 393 enumerators with all three kinds of reinterview sample cases in the reinterview program. (YYY) Source: d191.mastssn

| Rand  Admin  Supp<br>Case | 10 or more cases<br>selected for<br>reinterview | Fewer than 10<br>cases selected for<br>reinterive | Total   |
|---------------------------|---|---|---------|
| NNY                       | 7   | 16  | 23      |
| NYN                       | 49  | 183   | 232     |
| NYY                       | 1   | 0   | 1       |
| YNN                       | 2,299   | 438,707   | 441,006 |
| YNY                       | 6,212   | 2,200   | 8,412   |
| YYN                       | 13,648  | 2,054   | 15,702  |
| YYY                       | 389   | 4   | 393     |
| Total                     | 22,605  | 443,164   | 465,769 |

The SAS System  
The STEPDISC Procedure

11:45 Monday, May 12, 2003 4

The Method for Selecting Variables is STEPWISE

Observations 148238 Variable(s) in the Analysis 18  
Class Levels 2 Variable(s) will be Included 0  
Significance Level to Enter 0.15  
Significance Level to Stay 0.15

Class Level Information

| Variable Name | Frequency | Height | Proportion |
|---------------|-----------|--------|------------|
| 0             | 131922    | 131922 | 0.889934   |
| 1             | 16316     | 16316  | 0.110066   |

The SAS System

11:45 Monday, May 12, 2003 7

The STEPDISC Procedure

Stepwise Selection Summary

| Step | Number In | Entered  | Removed | Partial R-Square | F Value | Pr > F | Wilks' Lambda | Pr < Lambda | Average Squared Canonical Correlation | Pr > ASCC |
|------|-----------|----------|---------|------------------|---------|--------|---------------|-------------|---------------------------------------|-----------|
| 1    | 1         | hiscor   |         | 0.0009           | 130.62  | <.0001 | 0.99911963    | <.0001      | 0.00088037                            | <.0001    |
| 2    | 2         | izdel    |         | 0.0007           | 97.16   | <.0001 | 0.99846521    | <.0001      | 0.00153479                            | <.0001    |
| 3    | 3         | edulev   |         | 0.0001           | 10.04   | 0.0015 | 0.99839760    | <.0001      | 0.00160240                            | <.0001    |
| 4    | 4         | xdelscr  |         | 0.0000           | 4.99    | 0.0255 | 0.99836399    | <.0001      | 0.00163601                            | <.0001    |
| 5    | 5         | othlang  |         | 0.0000           | 4.65    | 0.0310 | 0.99833866    | <.0001      | 0.00166734                            | <.0001    |
| 6    | 6         | izdelva  |         | 0.0000           | 4.60    | 0.0319 | 0.99830166    | <.0001      | 0.00169834                            | <.0001    |
| 7    | 7         | tralev   |         | 0.0000           | 4.36    | 0.0367 | 0.99827227    | <.0001      | 0.00172773                            | <.0001    |
| 8    | 8         | priorexp |         | 0.0000           | 4.10    | 0.0428 | 0.99824464    | <.0001      | 0.00175536                            | <.0001    |
| 9    | 9         | xdeled   |         | 0.0000           | 3.24    | 0.0717 | 0.99822281    | <.0001      | 0.00177719                            | <.0001    |
| 10   | 10        | izdelpop |         | 0.0000           | 2.37    | 0.1239 | 0.99820685    | <.0001      | 0.00179315                            | <.0001    |
| 11   | 11        | izpopl   |         | 0.0000           | 4.59    | 0.0321 | 0.99817691    | <.0001      | 0.00182403                            | <.0001    |

